

BAB II

LANDASAN TEORI

2.1. Studi Pustaka

2.1.1. Data

Menurut Hoffer *et al.* (2004) data adalah segala sesuatu yang mewakili hal-hal dan peristiwa-peristiwa yang dianggap penting dan bermakna oleh pengguna (*user*).

Menurut Elmasri dan Navathe (2004) data adalah fakta yang dapat dicatat dan memiliki arti.

Dapat disimpulkan dari kedua definisi diatas bahwa data adalah fakta yang dapat dicatat, mewakilkan suatu objek serta memiliki arti yang penting bagi pengguna (*user*).

Pembagian tahap pengolahan data menurut Hartono (1999) terbagi menjadi 3 tahapan yang dikenal sebagai "siklus pengolahan data". Siklus tersebut yaitu:

1. Tahap *Input*

Data dimasukkan ke dalam komputer melalui alat input, seperti *keyboard*, *mouse*, atau perangkat lainnya.

2. Tahap *Processing*

Langkah berikutnya adalah pengolahan data yang telah dimasukkan pada tahap input. Proses ini dilakukan oleh perangkat pemroses, yang dapat melakukan perhitungan, perbandingan, kontrol, atau pencarian dalam penyimpanan.

3. Tahap *Output*

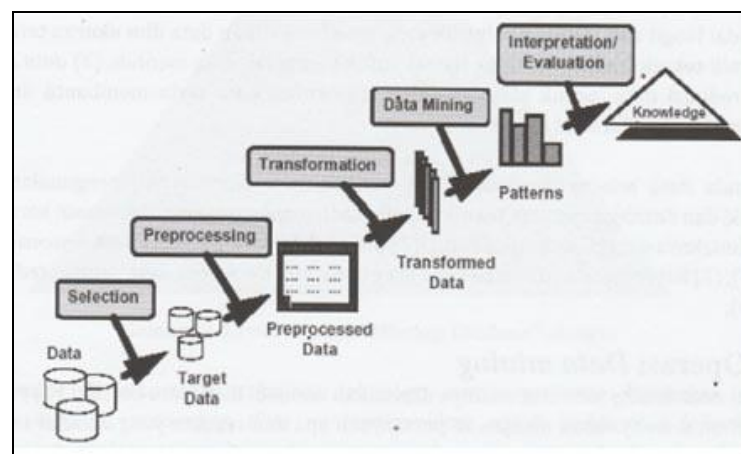
Perangkat output dihasilkan dari data yang telah diproses oleh proses, yang berupa informasi.

2.1.2. Data Mining

Menurut Davies dan Paul Beyton (2004), data *mining* merupakan proses penemuan atau penambangan informasi baru dengan mencari pola dari kumpulan data yang besar. *Discovery of knowledge in database* (KDD) adalah istilah tambahan untuk data *mining*, yang melibatkan penggunaan data historis untuk mengidentifikasi pola dan keteraturan dalam data berukuran besar. Setiap pola yang ditemukan diharapkan memberikan manfaat atau keuntungan tertentu. Proses ini dapat berlangsung secara otomatis atau semi-otomatis.

Hermawati (2013) menjelaskan bahwa data *mining* adalah proses yang menggunakan pembelajaran mesin untuk menganalisis dan mengekstraksi data secara otomatis.

A. Tahapan dalam Data Mining



Gambar 2.1 Tahapan Data Mining
(Sumber : Fayyad, 1996)

- *Selection*: Data yang diperoleh disaring terlebih dahulu. Melalui seleksi data, proses pengolahan dapat ditingkatkan untuk mencapai tujuan penelitian dengan lebih efektif.
- *Preprocessing*: serangkaian proses persiapan data sebelum data *mining* dilakukan. Tahapan *preprocessing* biasanya meliputi pembersihan (*cleaning*), pengurangan (*reduction*), dan integrasi data.

- *Transformation*: Data perlu mengalami transformasi sebelum diolah menggunakan data *mining*. Tujuannya adalah untuk menyesuaikan data yang akan diproses dengan algoritma dan perangkat lunak yang dipakai untuk mengolah data.
- *Data Mining*: Data diproses berdasarkan algoritma sesuai dengan teknik data *mining*. Dalam penelitian ini algoritma *naive bayes* digunakan, ini adalah salah satu algoritma dalam metode klasifikasi, dan outputnya dapat berupa prediksi atau klasifikasi berdasarkan data yang telah diproses.
- *Interpretation/evaluation*: Tahapan di mana hasil yang diperoleh dari teknik data *mining* diinterpretasikan.

B. Fungsi dalam Data Mining

Menurut Maclennan *et al.* (2009), data *mining* memiliki banyak fitur yang dapat digunakan dan dalam beberapa kasus, fitur/fungsi ini dapat membantu menyelesaikan masalah yang sedang dihadapi. Berikut fungsi data *mining* secara umum:

- *Klasifikasi (Classification)*
Fungsi dari klasifikasi adalah untuk mengelompokkan kelas berdasarkan kategori yang telah ditentukan. Algoritma yang digunakan antara lain: pohon keputusan, *K-nearest neighbor*, *neural networks*, *naive bayes*, dan *support vector machines*.
- *Clustering*
Clustering berfungsi untuk mencari atribut yang mempunyai kemiripan dan kemudian dikelompokkan kedalam segmen-segmen. Untuk proses *clustering*, algoritma yang dapat digunakan adalah *K-Means Clustering*, *Density-Based Clustering*, *Hierarchical Clustering*, dan *Means Shift Clustering*.

➤ *Association*

Association digunakan untuk menemukan hubungan antara fitur atau kumpulan item yang berasal dari *rule association* yang ada berdasarkan banyak item . Algoritma yang bisa digunakan adalah *FOLD Growth*, *Apriori*, *FP-Growth* dan .

➤ *Regression*

Seperti klasifikasi, tujuan regresi adalah menemukan prediksi dari pola yang sudah ada.

➤ *Forecasting*

Forecasting bertujuan untuk menghasilkan perkiraan waktu yang didasarkan pada tren dari data dahulu.

➤ *Sequence Analysis*

Sequence Analysis bertujuan untuk mengidentifikasi pola yang mungkin tersembunyi dalam urutan serangkaian peristiwa.

➤ *Deviation Analysis*

Deviation Analysis memiliki fungsi untuk menghasilkan kejadian yang berbeda dari kejadian biasa.

2.1.3. Klasifikasi

Menurut Han dan Kamber (2006), klasifikasi adalah metode untuk mengembangkan model atau fungsi yang menerapkan model matematika pada data atau konsep untuk membuat inferensi statistik untuk data yang tidak diketahui.

Menurut Kusriani *et al.* (2009) Variabel kategori digunakan dalam klasifikasi target. Sebagai contoh, klasifikasi keuntungan dapat dibagi menjadi tiga kategori: kecil, menengah, dan besar. Contoh klasifikasi bisnis dan penelitian lainnya adalah:

- Menentukan kecurangan dalam transaksi kartu kredit.
- Mempertimbangkan kredibilitas pengajuan hipotek klien.
- Mengidentifikasi jenis penyakit yang diderita pasien serta kategori penyakitnya.

2.1.4. Algoritma *Naive Bayes*

Menurut Bramer (2007) dalam buku yang berjudul “*Principles of Data Mining*” Algoritma *naive bayes* tidak memiliki aturan, metode ini menggunakan suatu cabang matematik yaitu konsep probabilitas. Konsep ini digunakan untuk menemukan probabilitas tertinggi dari proses klasifikasi yang dilakukan, dengan melakukan pengamatan terhadap kemunculan setiap klasifikasi dalam data pelatihan. Algoritma ini adalah salah satu yang populer yang dapat digunakan dalam pada penambangan data.

Dalam Bustami (2014) rumus teorema Bayes adalah :

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)}$$

Persamaan 2.1

Penjelasan dari persamaan 1 sebagai berikut:

X : Data yang memiliki kelas yang tidak diketahui.

H : Hipotesis untuk data X yang termasuk dalam kelas tertentu.

P(H|X) : Probabilitas nilai hipotesis H berdasarkan kondisi X.

P(H) : Nilai probabilitas hipotesis H.

P(X|H) : Probabilitas nilai hipotesis X berdasarkan kondisi H.

P(X) : Nilai probabilitas hipotesis X

a. Perhitungan *Prior*

Perhitungan yang dilakukan dengan menghitung perbandingan jumlah anggota kelas dengan seluruh sampel data.

$$P = \frac{X}{A}$$

Persamaan 2.2

Keterangan :

P : Angka *prior*

X : Total data untuk setiap kelas

A : Kumpulan data untuk semua kelas

b. Perhitungan *Likelihood*

Proses menghitung nilai probabilitas/kemungkinan untuk setiap kelas probabilitas kemunculan karakteristik dari karakteristik yang dipilih.

$$L = \frac{F}{B}$$

Persamaan 2.3

Keterangan :

L : Nilai probabilitas

F : Total data atribut untuk setiap kelas

B : Total seluruh data tiap kelas

c. Perhitungan *Posterior*

Hasil dari perhitungan *likelihood* berupa probabilitas atribut untuk suatu kelas, yang digunakan untuk menentukan probabilitas bahwa masing-masing atribut dimasukkan ke dalam kelas. Perhitungan kembali dihitung dengan mengalikan kemungkinan karakteristik input dengan kuadrat, dalam proses ini diperoleh probabilitas akhir dari hasil hasil tugas.

$$P(c) \pi P(a|c)$$

Persamaan 2.4

Keterangan :

P(C) : Angka *prior* tiap kelas

P(a|c) : Angka probabilitas

Hasil dari proses klasifikasi dibuat dengan melakukan perbandingan nilai *posterior* dari *class* yang ada. Angka *posterior* tertinggi adalah nilai yang dipilih dari klasifikasi.

2.1.5. *Confusion Matrix*

Menurut Gorunescu (2011) dalam buku berjudul “*Data Mining: Concepts, Models, and Techniques*”, *Confusion Matrix* merupakan alat visualisasi umum dalam pembelajaran yang diawasi. Setiap kolom dalam matriks mewakili turunan dari kelas predikat, dan setiap baris mewakili turunan aktual dari kelas tersebut. Kinerja sistem biasanya dievaluasi menggunakan matriks data yang tersedia. Berikut tabel 2.1 dibawah ini adalah contoh tabel untuk *confusion matrix*.

Tabel 2.1 *Confusion Matrix*

		<i>Predicted Sample</i>	
		<i>True</i>	<i>False</i>
<i>Actual Sample</i>	<i>True</i>	TP	FP
	<i>False</i>	FN	TN

Keterangan :

TP : *True positives*, banyaknya data dengan kelas positif tergolong positif.

TN : *True negative*, banyaknya data dengan kelas negatif tergolong negatif.

FP : *False positives*, total data dengan kelas positif tergolong negatif.

FN : *False negative*, banyaknya data dengan kelas negatif yang tergolong positif.

Dalam Han *et al* (2012), Efisiensi matriks konfusi adalah ukuran akurasi, presisi, dan nilai memori dari suatu model algoritma. Nilai akurasi adalah persentase akurasi antara nilai perkiraan dan nilai yang sebenarnya. Nilai presisi adalah presisi dari kelas yang diprediksi. Nilai *recall* adalah persentase dari nilai keberhasilan algoritma yang digunakan.

Nilai akurasi merupakan rasio dari jumlah total perkiraan yang benar, nilai akurasi yang tinggi dapat memperkirakan kebenaran lebih banyak, dapat dilakukan perhitungan menggunakan rumus berikut:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Persamaan 2.5

Recall dihitung untuk menilai kemampuan model dalam memprediksi kelas tertentu. Ini persis rasio antara jumlah data di kelas tertentu dibagi dengan jumlah semua kelas.

$$Recall = \frac{TP}{TP + FN}$$

Persamaan 2.6

Precision digunakan untuk menilai seberapa baik model dapat memperkirakan class. Akurasi ditentukan dengan menghitung rasio nilai data untuk kelas tertentu dibagi dengan jumlah total prediksi untuk kelas tersebut.

$$Precision = \frac{TP}{TP + FP}$$

Persamaan 2.7

Dalam Gorunescu (2011) nilai akurasi dapat dinilai berdasarkan beberapa kategori yang dapat dilihat dalam pada Tabel 2.2

Tabel 2.2 Kategori Nilai AUC

<u>Nilai AUC</u>	<u>Kategori Klasifikasi</u>
0.90-1.00	<i>Excellent Classification</i>
0.80-0.90	<i>Good Classification</i>
0.70-0.80	<i>Fair Classification</i>
0.60-0.70	<i>Poor Classification</i>
0.50-0.60	<i>Failure</i>

Menurut Martin dalam Lancaster (2019) untuk menginterpretasikan angka presisi dan *recall* dikategorikan menjadi tiga kelas. Dalam hal ini, metode yang digunakan adalah menggunakan skala interval dengan menghitung perbedaan antara nilai presisi dan *recall* tertinggi (1) dan nilai presisi dan *recall* terendah (0), kemudian membaginya menjadi tiga kategori penilaian. Tiga kategori dalam *precision* yang dapat dilihat pada Tabel 2.3

Tabel 2.3 Kategori Precision dan Recall

<u>Nilai Precision dan Recall</u>	<u>Kategori Klasifikasi</u>
0,67-1,00	<u>Tinggi</u>
0,34-0,66	<u>Sedang</u>
0,00-0,33	<u>Rendah</u>

2.1.6. Produk

Menurut Tjiptono (2008), produk adalah segala barang yang dapat disediakan oleh produsen untuk dipertimbangkan, dipakai, digunakan, dibuang, atau dikonsumsi oleh individu atau kelompok sebagai sarana untuk memenuhi kebutuhan atau memenuhi keinginan mereka.

2.1.7. Penjualan

Menurut Kotler (2008) penjualan adalah suatu proses sosial dimana individu dan kelompok memenuhi kebutuhan dan keinginan mereka serta

menciptakan, menyediakan, dan memasarkan barang dan jasa yang memiliki nilai dalam hubungannya dengan individu atau kelompok lain.

2.1.8. RapidMiner

Menurut C Aprilla *et al* (2013) dalam buku "Belajar Data Mining dengan RapidMiner", Analisis untuk data mining, text mining, dan prediksi disediakan oleh perangkat lunak *open source rapidminer*. Banyak metode deskriptif dan prediktif membantu orang membuat keputusan yang lebih baik. Kurang dari separuh operator penambangan data yang tersedia di *rapidminer* ditujukan untuk *input*, *output*, pra-pemrosesan, dan visualisasi data. Pada penelitian ini menggunakan versi 10.3 dari aplikasi *rapidminer*. Aplikasi *rapidminer* juga menyediakan fitur operator *naive bayes* yang bisa digunakan untuk mengolah data sesuai dengan algoritma dalam penelitian ini yaitu algoritma *naive bayes*.

2.2. Studi Literatur

2.2.1. Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Voucher Kuota di Edi Cell

H. Aini *et al.* (2023) menyatakan bahwa dalam penelitian ini membahas tentang penggunaan algoritma *naive bayes* untuk mengklasifikasikan produk yang paling banyak dijual pada penjual voucher kuota di Edi Cell. Penelitian ini didasari oleh kemajuan teknologi informasi dan komunikasi yang terus berlangsung dengan cepat dan harga yang semakin terjangkau, sehingga membuat Voucher Kuota menjadi kebutuhan yang mendesak bagi masyarakat saat ini. Beberapa produk kupon yang tersedia di EDI Cell antara lain voucher telkomsel, voucher XL, voucher Indosat, dan voucher 3. Produk-produk tersebut dikategorikan berkualitas rendah atau berkualitas tinggi. Tujuannya adalah untuk mempelajari bagaimana penggunaan algoritma naif bayes dalam data mining untuk menentukan klasifikasi produk berbasis kredit dan keakuratan data yang diperoleh. Berdasarkan penelitian, produk dengan nama produk voucher kuota tri adalah yang paling laris. Metode *naive bayes* pada penelitian ini menghasilkan skor *accuracy* 97,50%, *precision*

100%, dan skor *recall* 93,48%, yang menunjukkan bahwa metode ini cukup baik untuk penelitian ini.

2.2.2. Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat

Herry dan Saruni (2020) menyatakan bahwa dalam penelitian ini membahas tentang implementasi data *mining* pada penjualan obat menggunakan algoritma *naive bayes*. Dalam pengumpulan informasi mengenai produk-produk yang tersedia di apotek, misalnya vitamin, seringkali sulit untuk menentukan produk mana yang tersedia di apotek dan mana yang tidak. Hal ini karena suplemen vitamin memiliki umur simpan yang sedikit berbeda dengan vitamin yang tersedia di apotek. Untuk mengatasi permasalahan diatas maka obat-obatan yang dijual di apotek diklasifikasi berdasarkan variabel-variabel yang diperoleh. Dalam penelitian ini, alat *rapidminer* versi 8 digunakan untuk menguji data yang akan diolah untuk menghasilkan hasil akurasi dan nilai *ROC*. *Accuracy* hasilnya mencapai 88.00%.

2.2.3. Penentuan Tingkat Penjualan Mobil di Indonesia Dengan Menggunakan Algoritma Naïve Bayes

Romli *et al* (2019) menyatakan bahwa dalam penelitian ini membahas tentang penerapan algoritma *naive bayes* untuk menghitung tingkat penjualan mobil di Indonesia. Ada banyak merek mobil yang masuk ke pasar otomotif, tetapi GAIKINDO sendiri tidak memberikan klasifikasi pembeda untuk merek mobil yang paling diminati dan paling laris di situs webnya. Oleh karena itu, klasifikasi yang berbeda diperlukan agar pelanggan dapat memilih merek mobil mana yang paling laris berdasarkan kategori yang paling diminati. Selain itu, mendorong para produsen mobil untuk membuat mobil sesuai dengan keinginan pelanggan. Hasil penelitian memberikan kategori dan klasifikasi pembeda merek mobil yang paling diminati oleh pelanggan. Dengan nilai akurasi 92,19%, presisi 98,39%, dan nilai *recall* 87,14%, Metode *naive bayes* dianggap bisa digunakan dalam penelitian ini.

2.2.4. Penerapan Data Mining Dengan Metode Naive Bayes Classifier Pada Penjualan Barang Untuk Optimasi Strategi Pemasaran

Lubis dan Sitohang (2023) menyatakan bahwa dalam penelitian ini membahas tentang peningkatan strategi pemasaran menggunakan data *mining* dengan teknik klasifikasi *naive bayes*. Salah satu masalah yang dihadapi oleh toko ini adalah data transaksi penjualan hanya disimpan, yang berarti mereka mengumpulkan data tanpa mengetahui manfaatnya. Toko *Universal Homeware* saat ini tidak tahu bagaimana mengolah data untuk memahami pertumbuhan bisnis atau operasi yang sedang berlangsung. Selain itu, ketika data tidak tersedia, sering terjadi pembelian barang yang dibutuhkan pelanggan. Hasil analisis metode yang menggunakan algoritma *naive bayes* dengan 36 titik data menghasilkan akurasi sekitar 80%. Tercatat pada tahun 2020, 38% transaksi produk laris dan 30% produk tidak laris. Program *rapidminer* dapat membantu dalam memberikan bukti bahwa hasil pengujian dengan metode *naive bayes* menghasilkan nilai akurasi yang baik.

2.2.5. Penerapan Data Mining Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Smartphone Menggunakan Metode Naive Bayes

Fauzi (2022) menyatakan bahwa dalam penelitian ini membahas tentang manfaat metode data *mining naive bayes* untuk mengklasifikasikan produk *smartphone*. Penelitian ini didasarkan pada kemajuan teknologi yang pesat dan biaya teknologi komunikasi dan informasi yang cukup terjangkau. Kemajuan ini membuat setiap orang di dunia saat ini harus memiliki *smartphone*. *Smartphone Samsung, iPhone, Vivo, dan Oppo* termasuk dalam daftar produk yang tersedia di toko *Siantar Store*. Peneliti mengkategorikan produk menjadi laris dan tidak laris. Tujuannya adalah untuk mengetahui bagaimana data *mining* dapat digunakan untuk mengidentifikasi klasifikasi produk yang paling laris dan hasil akurasi data tentang stok penjualan *smartphone*. Berdasarkan hasil penelitian, *smartphone* merek *Samsung* termasuk dalam kategori produk premium di pasar *smartphone*,

dengan akurasi klasifikasi sebesar 97,50% menggunakan *metode naive bayes*, presisi sebesar 100%, dan *recall* sebesar 94,38%.

2.2.6. Menentukan Menu Makanan Favorit di Outlet Barbar Sampit Menggunakan Algoritma Naive Bayes

Emawati *et al* (2022) menyatakan bahwa dalam penelitian ini membahas tentang penggunaan algoritma *naive bayes* untuk mengidentifikasi menu makanan favorit di outlet Barbar sampit. Makanan korea seperti sayap ayam yang dicampur dengan saus seperti saus pedas atau manis disajikan di Outlet Barbar. Ketika ingin melakukan prediksi, mengumpulkan daftar makanan apa yang paling sering dibeli adalah salah satu masalah yang sering terjadi. Metode ini sangat sulit karena banyaknya data yang dicatat dalam setiap transaksi. Akibatnya, diperlukan metode yang akan memudahkan pelanggan untuk memprediksi apa yang akan ada di menu. Hasil prediksi menunjukkan bahwa pada bulan Juni, menu paling favorit adalah 20 Winger Komplit dengan persentase 15,01%; pada bulan Juli adalah 25 Winger Komplit dengan persentase 17,06%; dan pada bulan Agustus adalah 25 Winger Komplit dengan persentase 19,36%..

2.2.7. Algoritma Naïve Bayes Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar

Juwita *et al* (2022) menyatakan bahwa dalam penelitian ini membahas tentang algoritma *naive bayes* untuk memperkirakan penjualan di toko VJCakes Pematang Siantar. Penjualan produk kue diperkirakan akan terus mendongkrak pendapatan. Salah satu caranya adalah dengan menawarkan barang sesuai dengan permintaan konsumen untuk mencegah kerugian. Hingga saat ini, banyak pelaku bisnis yang sering mengalami kerugian akibat tidak adanya sistem prediksi penjualan. Tujuan artikel ini adalah untuk menerapkan dan mendemonstrasikan penerapan algoritma *naive bayes* untuk memperkirakan penjualan kue di toko VJCakes Pematang Siantar. Temuan yang dihitung menunjukkan bahwa metode perhitungan manual dan metode bantuan perangkat lunak *rapidminer* memberikan

hasil yang sama. Hal ini menunjukkan bahwa perhitungan tersebut efektif karena menghasilkan tingkat akurasi sebesar 83,44% dan tabel probabilitas masing-masing variabel.

2.2.8. Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat

Harahap *et al* (2021) menyatakan bahwa dalam penelitian ini membahas tentang penggunaan data *mining* dan algoritma klasifikasi *naive bayes* untuk memprediksi pembelian cat. Permasalahan yang diteliti adalah bagaimana agen pembeli menciptakan rasa urgensi pada saat membeli cat dari pemasok dengan cara mempererat hubungan antara pembeli dan sponsor, sehingga produk tersebut menjadi kurang diminati pelanggan dan akhirnya berpindah ke yang lain. perusahaan yang menjual cat yang lebih sesuai dengan kebutuhannya. Algoritma *naive bayes* berhasil mengklasifikasikan 48 dari 60 titik data menggunakan hasil proses pelatihan data cat yang diubah menjadi data pelatihan, dan mencapai tingkat akurasi sekitar 80%..

2.2.9. Klasifikasi Penjualan Produk Menggunakan Algoritma Naive Bayes pada Konter HP Bayu Cell

Ayu dan Putra (2024) menyatakan bahwa dalam penelitian ini membahas tentang algoritma *naive bayes* yang digunakan di konter HP Bayu Cell untuk mengkategorikan penjualan produk. Konter HP Bayu Cell menghadapi permasalahan untuk menentukan produk apa yang laris dan kurang laris. Untuk mengatasi masalah ini dan mencegah penumpukan produk yang tidak terjual, penelitian ini menggunakan algoritma data *mining*, yaitu algoritma *naive bayes*. Algoritma ini memanfaatkan peluang suatu kejadian berdasarkan kejadian lain yang telah terjadi. Pada penelitian ini, dilakukan analisis dengan algoritma *naive bayes* pada data *training* dengan fokus pada produk handphone REDMI NOTE 9. Hasil dari analisis menunjukkan bahwa produk tersebut diklasifikasikan ke dalam kategori "Laris" dengan nilai peluang sebesar 0,005211. Dengan demikian, sistem yang

diimplementasikan pada konter HP Bayu Cell telah memberikan hasil yang diharapkan, membantu dalam mengidentifikasi produk handphone mana yang masuk ke dalam kategori laris dan tidak laris.

2.2.10. Klasifikasi Data Penjualan Alat Tulis Kantor (ATK) Terlaris Untuk Optimasi Strategi Pemasaran Di Toko Citra Media Menggunakan Metode *Naive Bayes*

Widodo dan Sa'adah (2019) menyatakan bahwa dalam penelitian ini membahas tentang penggunaan metode *naive bayes* untuk mengklasifikasikan data penjualan ATK terlaris sehingga toko citra media dapat mengoptimalkan strategi pemasarannya. Dari pengamatan langsung di toko citra media, terlihat bahwa pendataan barang di gudang sering mengalami kesalahan karena masih dilakukan secara manual dan tidak terstruktur. Data barang didata dengan buku dan kemudian dimasukkan satu per satu ke dalam *ms. excel* tanpa proses yang terorganisir, sehingga tidak semua barang terinput dengan lengkap ke dalam tabel *excel*. Penelitian ini bertujuan untuk membuat strategi pemasaran yang dapat membantu toko mengelola data stok barang dengan lebih baik, sehingga dapat lebih responsif terhadap kebutuhan pelanggan dan meningkatkan keuntungan penjualan serta mengetahui tren ataupun produk yang sedang diminati oleh pelanggan. Penelitian tersebut menghasilkan akurasi yang sangat tinggi, dengan akurasi tertinggi mencapai 99,26% dan akurasi terendah mencapai 94,39%. Dengan menggunakan kurva *AUC*, evaluasi menyatakan bahwa hasil dari delapan pengujian masuk ke dalam klasifikasi yang sangat baik, dengan batas atas sebesar 0,997. Hal ini menunjukkan bahwa sistem yang diuji telah berhasil dengan baik dalam mengklasifikasikan dan mengelola data stok barang dengan efisien dan akurat, memberikan potensi untuk meningkatkan kinerja toko dalam memenuhi kebutuhan pelanggan dan meningkatkan keuntungan penjualan.

Tabel 2.4 Penelitian Terdahulu

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
1.	Penerapan Algoritma <i>Naive Bayes</i> Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Voucher Kuota Di Edi Cell	2023	Nur Aini H, Muchlis, Deo Novaldi, Andestan Wirayuda, Ivan Mei Dwintara	Akibat pesatnya kemajuan teknologi informasi dan komunikasi, kini masyarakat sudah menganggap Voucher Kuota sudah menjadi suatu kebutuhan. Produk Voucher Kuota dari Telkomsel, XL, Indosat, dan tiga lainnya masuk dalam kategori laris dan kurang laris di EDI Cell.	Untuk memahami pemanfaatan penambangan data bersama dengan algoritma <i>naive bayes</i> untuk mengkategorikan barang terlaris dan menilai ketepatannya menggunakan data terkait kredit.	<i>Naive Bayes</i>	Menurut penelitian, produk yang dikenal sebagai voucher tri adalah yang paling laris. Algoritma <i>naive bayes</i> memperoleh skor 97,50%, recall 93,48%, dan presisi 100%, yang menunjukkan bahwa metode ini sangat efektif untuk penelitian.

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
2	Implementasi Data Mining dengan Algoritma <i>Naïve Bayes</i> pada Penjualan Obat	2020	Herry Derajad Wijaya dan Saruni Dwiasnati	Memberikan informasi tentang penjualan barang di sebuah apotek, terutama vitamin, seringkali menjadi tantangan karena sulit untuk menentukan barang mana yang diminati oleh konsumen. Situasi ini disebabkan oleh dua faktor utama: Pertama, terdapat akumulasi stok vitamin yang kurang diinginkan di gudang, kedua, vitamin yang dicari tidak tersedia di persediaan apotek	Tujuan dari penelitian ini adalah untuk menganalisis masalah-masalah yang terkait dengan penentuan apakah sebuah produk dapat dikategorikan sebagai laku atau tidak laku.	<i>Naive Bayes</i>	Dataset sebanyak 150 digunakan dalam penelitian ini. Studi ini dilakukan karena ada masalah untuk mengetahui stok produk pada bulan sebelumnya untuk menentukan stok mana yang harus diperbanyak lagi dan mana yang harus dikurangi. Untuk pengujian model data, seluruh set data digunakan untuk menguji data rekapitulasi penjualan obat, yang menghasilkan tingkat akurasi 88.00%.

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
3.	Penentuan Tingkat Penjualan Mobil Di Indonesia Dengan Menggunakan Algoritma <i>Naive Bayes</i>	2019	Ikhsan Romli, Esem Pusnawati, Arif Siswandi	Beragamnya merek mobil yang terlibat di pasar otomotif memerlukan identifikasi merek yang paling banyak dicari konsumen dan menentukan kategori terlaris. Namun situs GAIKINDO tidak memberikan informasi yang jelas mengenai hal ini.	Untuk memberikan informasi mengenai merek mobil terlaris yang paling banyak diminati oleh konsumen	<i>Naive Bayes</i>	Hasil penelitian menunjukkan bahwa terdapat klasifikasi yang membedakan merek mobil yang paling diminati oleh konsumen dan kategori yang paling laris. Dengan menggunakan Metode <i>naive bayes</i> , tingkat akurasi klasifikasi mencapai 92,19%, dengan nilai presisi sebesar 98,39% dan nilai <i>recall</i> sebesar 87,14%. Hal ini menunjukkan bahwa metode <i>naive bayes</i> berhasil memberikan hasil yang baik dalam penelitian ini, dengan tingkat keakuratan yang cukup tinggi.

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
4.	Penerapan Data Mining Dengan Metode <i>Naive Bayes Classifier</i> Pada Penjualan Barang Untuk Optimasi Strategi Pemasaran	2023	Kahfi Rama Putra Lubis, Sunarsan Sitohang	Salah satu masalah yang dihadapi oleh penelitian ini adalah bahwa data transaksi penjualan hanya disimpan, yang berarti mereka mengumpulkan data tanpa mengetahui manfaatnya. Toko <i>Universal Homeware</i> saat ini tidak tahu bagaimana mengolah data Sangat penting untuk terus mendapatkan informasi terkini tentang aktivitas dan perkembangan bisnis saat ini. Selain itu,	Untuk mengidentifikasi keuntungan dari data transaksi yang dikumpulkan dari toko tersebut untuk mengetahui perkembangan bisnis atau bisnis yang sedang dijalankan.	<i>Naive Bayes</i>	<i>Software rapidminer</i> dapat membantu menunjukkan bahwa perhitungan algoritma <i>naive bayes</i> menghasilkan nilai akurasi yang tinggi. Hasil analisis metode yang diuji dengan 36 dataset menunjukkan 30% penjualan berhasil dan 38% tidak menghasilkan penjualan pada tahun 2020.

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
				pengadaan barang yang dibutuhkan pelanggan sering tidak ada.			
5	Penerapan Data Mining Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan <i>Smartphone</i> Menggunakan Metode <i>Naive Bayes</i>	2022	Rival Fauzi	Dengan kemajuan pesat dan terjangkau dalam teknologi komunikasi dan informasi, memiliki <i>smartphone</i> sekarang menjadi kebutuhan. Dengan mempertimbangkan berbagai produk yang tersedia di toko Siantar Store, termasuk <i>smartphone</i> Samsung, <i>iPHONE</i> , Vivo, dan Oppo, peneliti menentukan	Untuk mengetahui bagaimana data <i>mining</i> digunakan untuk mengidentifikasi kategorisasi produk terlaris dan nilai keakuratan data mengenai inventaris penjualan <i>smartphone</i> .	<i>Naive Bayes</i>	Pada penelitian ini algoritma <i>naive bayes</i> dinilai sangat efektif dengan tingkat akurasi klasifikasi sebesar 97,5%, nilai presisi sebesar 100%, dan nilai <i>recall</i> sebesar 93,48%.

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
				kategori produk menjadi laris dan tidak laris.			
6	Menentukan Menu Makanan Favorit di Outlet Barbar Sampit Menggunakan Algoritma <i>Naive Bayes</i>	2022	Nindi Emawati, Mohammad Nufan Nur Alif, Kanta Pramuda	Salah satu tantangan umum dalam membuat prediksi adalah mengakses catatan transaksi penjualan dan menyusun daftar barang yang sering dibeli. Tugas ini seringkali menantang karena banyaknya volume data transaksi yang terlibat.	Untuk memberikan informasi tentang produk terlaris dan tidak laris untuk evaluasi outlet	<i>Naive Bayes</i>	Hasilnya adalah Winger Komplit 20 sebagai Menu Makanan Favorit pada bulan Juni dengan nilai persentase 15,01%, Winger Komplit 25 sebagai Menu Makanan Favorit pada bulan Juli dengan nilai persentase 17,06%, dan Winger Komplit 25 sebagai Menu Makanan Favorit pada bulan Agustus dengan nilai persentase 19,36%.
7	Algoritma <i>Naive Bayes</i> Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar	2022	Juwita, M. Safii, Bahrudi Efendi Damanik	Produk kue diharapkan tetap menguntungkan dari segi penjualan, salah satunya dengan memastikan bahwa produk sesuai dengan permintaan	Untuk mengimplementasikan dan membuktikan bahwa Algoritma <i>naive bayes</i> dapat dipakai untuk memperkirakan penjualan kue di	<i>Naive Bayes</i>	Hasil perhitungan menunjukkan bahwa proses perhitungan manual dan yang dibantu oleh <i>software RapidMiner</i> adalah sama, Hal ini menunjukkan bahwa perhitungan dapat dianggap berhasil.

Tabel 2.4 Penelitian Terdahulu (Lanjutan)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
				pasar untuk mencegah kerugian. Pada periode tersebut, perusahaan seringkali mengalami kerugian karena tidak adanya sistem penjualan yang dapat diandalkan.	toko VJCakes Pematangsiantar.		Hasilnya menunjukkan tabel probabilitas untuk variabel dan akurasi sebesar 83,44% untuk uji coba data.
8	Penerapan Data Mining Dengan Algoritma <i>Naive Bayes Classifier</i> Dalam Memprediksi Pembelian Cat	2021	Fitriana Harahap, Nidia Enjelita Saragih, Elida Tuti Siregar, Husin Sariangshah	Salah satu masalah yang muncul pada manajer pembelian adalah ketika mereka merasa tertekan untuk membeli kucing melalui pemasok berdasarkan beberapa koneksi dekat dengan pemasok dan sejumlah biaya	Untuk menentukan jenis cat yang dibeli konsumen	<i>Naive Bayes</i>	Hasil penelitian menunjukkan keberhasilan metode <i>naive bayes</i> dalam memprediksi pembelian cat, mencapai tingkat akurasi 80% dengan mengklasifikasikan 48 dari 60 data yang diuji secara tepat.

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
				<p>sponsor yang signifikan yang diberikan kepada bisnis tersebut. Benar sekali, hal ini akan mendorong pelanggan untuk mengunjungi bisnis tersebut guna mendapatkan perlengkapan makan yang paling penting.</p>			
9	<p>Klasifikasi Penjualan Produk Menggunakan Algoritma <i>Naive Bayes</i> pada Konter HP Bayu Cell</p>	2024	Pradani Ayu Widya Purnama dan Teri Ade Putra	<p>Konter HP Bayu Cell menghadapi tantangan dalam mengidentifikasi produk mana yang berkinerja baik dan mana yang tidak. Untuk menghindari penumpukan produk, Konter HP</p>	<p>Untuk mencari produk HP pada Bayu Cell yang paling diminati</p>	<i>Naive Bayes</i>	<p>Dengan menggunakan algoritma <i>Naive Bayes</i> pada data latih menggunakan smartphone REDMI CATATAN 9, klasifikasi menghasilkan kategori “Laris” dengan nilai 0,005211. Oleh karena itu, akun HP Bayu kini dapat memberikan hasil yang diharapkan dan</p>

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
				Bayu Cell perlu memiliki sistem yang membantu mereka membuat keputusan dengan cepat dan tepat.			membantu mengidentifikasi produk ponsel cerdas sebenarnya, termasuk yang tergolong laris dan tidak laris.
10	Klasifikasi Data Penjualan Alat Tulis Kantor (Atk) Terlaris Untuk Optimasi Strategi Pemasaran Di Toko Citramedia Menggunakan Metode <i>Naive Bayes</i>	2019	Edy Widodo dan Pathyatus Sa'adah	Menurut observasi langsung yang dilakukan oleh peneliti di Toko Citra Media, karena pendataan barang di gudang toko masih dilakukan secara manual, terdapat kesalahan dalam pencatatan data dalam satu buku. dan kemudian dimasukkan ke dalam <i>Ms. Excel</i> secara manual, namun tidak dilakukan dengan teratur, sehingga tidak semua barang	Tujuan penelitian ini adalah merumuskan strategi pemasaran yang bertujuan untuk mengoptimalkan pengadaan barang di toko. Hal ini bertujuan untuk memastikan bahwa toko dapat memenuhi kebutuhan pelanggan dengan lebih baik dan meningkatkan keuntungan penjualan.	<i>Naive Bayes</i>	Penelitian tersebut menghasilkan akurasi yang sangat tinggi, Dengan akurasi tertinggi mencapai 99,26% dan akurasi terendah sebesar 94,39%, evaluasi tersebut menunjukkan kemampuan model. Memanfaatkan kurva <i>AUC</i> , penilaian menunjukkan bahwa hasil 8 pengujian masuk dalam kategori klasifikasi “sangat baik”, dengan nilai maksimum 0,997

Tabel 2.4 Penelitian Terdahulu (*Lanjutan*)

No.	Judul Penelitian	Tahun	Nama Peneliti	Permasalahan	Tujuan Penelitian	Metode/Model/Algoritma	Hasil Pembahasan
				tercatat secara lengkap dalam tabel <i>Excel</i> . Situasi ini berdampak pada proses penjualan.			